Review #1

The manuscript by Goldberg et al. is a valuable and timely analysis of NOx emissions during KORUS-AQ. It identifies some potential issues with NOx emissions in the region that are useful for air quality management as well as other works studying pollution during this campaign period. The work also has relevance beyond KORUS-AQ in terms of how OMI data is used to estimate NOx from urban areas, and also how TROPOMI data will be used in such studies in the future. The article is in general quite clear and easy to read, and most figures are useful and essential.

That being said, the work misses a critical opportunity to evaluate one of their main hypotheses, which is that regionally-derived NO2 columns (using air mass factors from high-resolution WRF-Chem simulations) lead to objectively better NOx inversions. In fact, while they report the difference between these NOx inversions and those based on the standard OMI NO2 data, the differences aren't critically evaluated, which is a shame, as it seems to be a rather easy next step. This would thus be my primary suggestion for revision. A few other aspects such as how using AMFs derived from a model that is clearly inaccurate to begin with affect their analysis, why spatial averaging is presented and then discarded, and why the regionally-derived NO2 columns may be overestimating NO2 in rural areas need to also be addressed.

Details of these comments as well as other are presented below; addressing them likely constitutes major revisions as additional WRF-Chem calculations are required.

Thank you for your comments; they have substantially improved our manuscript.

<u>Major comments:</u>

√Section 3.6: It isn't clear to me why the authors test a doubling of the emissions. The prior bottom-up values are 198, the top-down using standard product are 353 (an increase of x1.78) and the top-down using the regional product are 484 (an increase of x2.44). The test increase of x2 thus does little to distinguish between these two. This is a bit of a disappointment, as a major conclusion from this work is that the regional product (and top-down emissions using this product) are significantly different and better than the standard product. However, the only evidence presented that the regional product is better than standard thus far is the comparison to Pandora data. While encouraging, the authors are missing an big opportunity to make this argument much stronger by performing two model simulations for the entire KORUS-AQ period with top-down emissions that match those derived using the standard product and the regional product, precisely, and not some estimate of x2 that is neither here nor there. These two different model simulations can then be evaluated using the aircraft data.

In this revised manuscript, we have completed a month-long simulation with NOx emissions increased by a factor of 2.13, and have removed the two-day 2 × NOx scenario. A factor of 2.13 is chosen because the top-down estimate from the satellite is 484 kton/yr, while the top-down approach applied to the model is 227 kton/yr. The bottom-up NOx emissions inventory within a 40 km radius of Seoul is 198 kton/yr, however the 227 kton/yr value is a more appropriate comparison with the top-down satellite analysis.

We are confident the OMI-Regional NO_2 product is more robust than the standard product due to the comparison with the Pandora NO2 network. Furthermore, the methodology of updating the satellite product with high-resolution a priori NO_2 shape profiles is more scientifically appropriate for regional studies (Russell et al., 2012, Lamsal et al., 2015, Kuhlmann et al., 2016, Goldberg et. al., 2017).

Thus, we feel that it is unnecessary to perform a simulation with NOx increased by a factor of 1.56 (353 kton/yr vs. 227 kton/yr). Furthermore, as we show in two new figures, the updated 2.13 x NOx simulation agrees well with the aircraft data (Figure 9) and the OMI-Regional NO_2 product (Figure 10).

√General: Model values of NO2 column are much lower than regionally-derived OMI NO2 column in most areas, including rural areas (Fig 3). However model values match the aircraft data in rural areas (i.e. the only major discrepancies noted in discussion of Fig 5 or e.g. the conclusions (12.17-19)). What are we thus to make then of the quality of the regionally-derived OMI values in rural areas? Too high? This should be discussed. If these are too high, will the background values estimated in the EMG value thus be too high, and this error propagate into an error in the urban emissions?

In the original figure, we are referring to the "mainland transect". This is a subset of the rural areas, and was inappropriate. We have since updated the figure to include all mainland areas away from Seoul, and now find a discrepancy between NO_2 in the lowest layers between the model and the aircraft observations. This figure and corresponding discussion has been updated.

√General: If model columns are too low, how does that impact model calculated AMF? How much would AMF change if using posterior emissions in WRF-Chem? An additional calculation of AMFs based on WRF-Chem simulations with adjusted emissions needs to be performed to answer this question.

A new figure, Figure 11, now addresses this. The effect of the emissions inventory on the air mass factor is appreciable, but is secondary to the resolution of the model simulation. In the Seoul metropolitan area, the AMF changes on average by 35% when switching from GMI to WRF-Chem and changes by only 8% when switching emission inventories.

√Or perhaps the NO2 profiles in WRFChem are adjusted to account for this bias (this is indicated on 4.23, but no details are provided as to what this adjustment is, or how it is derived)? I try to evaluate the WRF-Chem profiles visually, based on Fig 5, but this plot doesn't make that information clearly visible given the way the vertical axis isn't strictly used (i.e. model and aircraft data collected at the same height are not plotted at the same height – which I understand from the perspective of clarity in showing their differences with box-whisker plots, but something else is needed to evaluate profile shapes).

The OMI-Regional NO_2 product derived herein already accounts for any mean model biases. A better description of this process is now provided in Section 2.1.1.

√General: if results with spatial ave kernel are not trusted for analysis, they should be removed throughout from the results. Otherwise, it is a bit of a distracting / potentially misleading presentation. For example on page 12, line 5 – this isn't used, so why is it highlighted here? Still, wouldn't there be some data from KORUS-AQ with which wind field estimates in WRF could be evaluated? It just seems a bit subjective here that this source of error is singled out (11.18) as justification for not using this approach, whereas profile shapes that come from WRF-Chem are deemed acceptable, even though WRF-Chem NO2 column values are significantly biased low in urban areas. Further, it seems that comparison to the Pandora data in Fig 6 would indicate that the spatial kernel adjustment is improving, rather than degrading, the column estimates, which is a point in favor of this approach.

As noted, the spatial averaging kernel provides important insight into resolving discrepancies between OMI NO_2 and Pandora NO_2. However, we also emphasize that the spatial averaging kernel has its limitations. The top-down approach is extremely sensitive to wind direction, so any errors in the forecasted wind fields will propagate through to the top-down method. When we apply a spatial averaging kernel to the satellite retrieval and then perform the top-down method, a NOx emissions rate cannot be derived. Therefore, for the top-down analysis, the artificial error introduced by spatial averaging kernel outweighs its benefits. However, for the Pandora comparison, the benefits outweigh the artificial errors (as shown in Figure 6).

√9.30-34: Not sure how this statement about NOx diurnal variability contributes to the difference between modeled and observed NO2 columns. Are the authors suggesting that the diurnal variability of NOx emissions in Korea is incorrect? Simply noting that it is different than modeled diurnal variability in the US is not sufficient evidence and in fact comes across as tangential, unless the authors are claiming that NOx source profiles (EGUs, distribution of diesel vehicles in the transportation fleet) are identical, which seems dubious. So I suggest removing Fig 4, unless this argument can be substantially strengthened.

We are suggesting that the temporalization of NOx emissions can introduce errors in satellite and aircraft measurements, which occur during the daytime. The temporalization is a best estimate based on literature, but it is almost certainly not correct either. The temporalization of NOx emissions as a major source of the discrepancy has not been discussed in previous literature and is quite critical to the conclusions of this manuscript. Resolving these differences is an important topic for future research.

However, we are not necessarily suggesting that the Korean temporalization is identical to the eastern US, but instead are providing a comparison to show how temporalization can differ by region.

The discussion of this topic in the text has been added to and is now referenced in the Conclusions as an important source of the discrepancy.

√Additionally, I wonder to what extent excessive NO2 deposition in the model might be contributing to the noted differences; this could be driven by e.g. PBL heights in the model that are too low. I suspect there is more information from the KORUS-AQ campaign that could be used to evaluate this.

We have now included a comparison with NO_y. Evaluating the NO_2 deposition rates and PBL heights is beyond the scope of this study.

√Fig 5 and associated text: I agree this suggests the differences between WRF-Chem and OMI near Seoul are likely driven by emissions, rather than chemistry, deposition, or PBL heights, as suggested by the authors or myself.

√10.20: Thoughts on why bias improves but not correlation? This might suggest that the daily variability of WRF-Chem (which impacts daily AMFs) is not correct, or at least not an improvement upon larger-scale averages.

Yes, these are our thoughts too. The WRF simulation used to drive the chemistry is in forecast mode. This has been clarified in the text.

√General: How does the plume analysis / rotation / EMG inversion process work if e.g. there is a large point source whose outflowing plume flows over another source (e.g. a highway) that runs parallel underneath it, replenishing NO2 concentrations that are then going to be ascribed only to emissions at a single point of plume origin? So, related, at 11.10: Yes, but the concern is rather smaller sources within this radius but not at the center that contribute to the plume (i.e. mobile sources).

Small sources at the edge of the urban boundary will lead to an artificially longer NO2 lifetime. This partially compensates the error introduced by the wind. A short commentary has now been included in the Section 3.6.2 of the manuscript.

<u>Minor comments and corrections:</u>

√Throughout: “shape profiles” reads a bit strange. Change to “profile shapes”? Or just profiles?

Updated

√1.25: for the –> for 1.26: “larger near large” rewrite

Updated

√2.4-5: “another . . . another” rewrite

Updated

√3.27: trace-gas Eq. 2: include a proper summation index

Updated

√4.5-6: It isn’t clear here if the authors are discussing how AMFs are calculated in general, in the standard retrieval, or in their own regionally-specific retrieval. Please clarify.

It is in reference to all OMI NO_2 products derived from the NASA OMI NO_2 product. This includes both the standard product and the regional product derived here (as well as any other custom products derived from the NASA product). It has been clarified.

√5.1: How big of an assumption is this, that the profiles are constant over this time range?

Please reference Laughner et al., 2016, which is already cited here. That study shows that the AMF can vary by 20% on a daily basis.

√6.26: I’m pretty sure AOD from geostationary satellites over Korea have been used for forecasting studies.

The sentence referring to this simulation as the first near real-time application of geostationary data has been removed.

√6.26: Not sure though how the authors here qualify their study as "nearreal time"; all I saw was reanalysis. NRT usually means forecasting. Just because the winds were forecast within the domain doesn't mean this is a chemical forecast, since the observations used span the time period over which the analysis (aircraft obs) are made (considerably, given that satellite data for several more years and months are used). This entire approach would be impossible in an NRT setting, given the data requirements for oversampling.

This statement is in reference to the model simulation only. The model simulation was indeed performed as a forecast in near-real time. The OMI NO_2 satellite data was processed after the fact, but AOD was in fact assimilatiated in near-real time.

√7.28: plume, –> plume

Updated

√8.6: Why using wind estimates from a different model than the one used to constrain WRF met at the boundary (NCEP), or different from WRF itself?

The WRF simulation is a forecast simulation. Re-analysis data is more robust despite it being at a coarser spatial resolution.

√8.8: Why 500m? Based on Fig 5 it looks like NO2 plumes extend much higher than that, up to 1 km or possibly above (although a bit hard to tell from this plot, given the manner in which the vertical scale is treated).

We follow Lu et al., 2015. Generally, winds do not vary much between 500 – 1000 m. De Foy et al., 2014 discuss how the selection of wind speeds/direction affect the top-down calculation. This is taken into account in the uncertainty analysis.

√Fig 1: content –> concentration

The word "content" is correct in this context. Concentration is mass per unit volume, which is not being shown here.

√Fig 1: Why showing US domain?

This has now been removed, but the US is still referenced in the text for comparison.

√9.4: is in despite of --> is despite

Updated

√Section 3.1: Inclusion of / comparison to the US feels tangential and unnecessary. Suggest focus on Korea domain; remove US from Fig 1 and remove discussion here. This point could be touched on in intro or conclusions but doesn't fit well in the results.

The US figure has now been removed, but the US is still quickly referenced in the text of this section for comparison.

√9.17: There are also small decreases in the southern part of the peninsula, as well the SE corner of the domain. Further, the explanation provided for the decreases isn't particularly insightful.

This sentence has been removed.

√9.21: From the presence of red in panel (c), the statement "in all areas" does not seem to accurately describe the results. Please update text to more precisely reflect the findings.

The word "all" has been changed to "most"

√Section 3.3.1: it's not good style to have only one subsub section in a section. Consider merging this with 3.3 or making 3.3 WRF-Chem evaluation, 3.3.1 comparison to OMI and 3.3.2 comparison to aircraft.

This section is now a section by itself, since it is now expanded.